CALIFORNIA
DATA
COLLABORATIVE

ATTN: Christina McCready, Department of Water Resources

CC: Greg Gearhard, State Water Resources Control Board
Stuart Drowns, Government Operations Agency
Angelica Quirarte, Government Operations Agency
Debbie Franco, Office of Planning and Research
George Isaac, Delta Stewardship Council
Tim Garza, Department of Water Resources
Kamyar Guivetchi, Department of Water Resources

We write to support the Department of Water Resources' (DWR) and other state agencies' efforts to develop a statewide water data platform and offer constructive comments to help achieve the goals of AB1755, the Open and Transparent Water Data Act (signed September 23, 2016). In particular, we would like to invite DWR to speak on AB1755 implementation at the next California Data Collaborative quarterly steering committee meeting during the afternoon of April 19, 2018 at Inland Empire Utilities Agency.

The California Data Collaborative (CaDC) is a coalition of local water utilities and in-kind academic and technology partners committed to sharing data into a common nonprofit platform (the Strategic California UrBan Analytics "SCUBA" data warehouse[1]) to support water managers into an uncertain future. The CaDC has been honored by the White House as part of its 2016 Water Summit, won best urban tool in the 2016 water data challenge, and has been featured as a leading water data initiative by numerous media outlets including Harvard Business Review, Harper's magazine and the Associated Press.

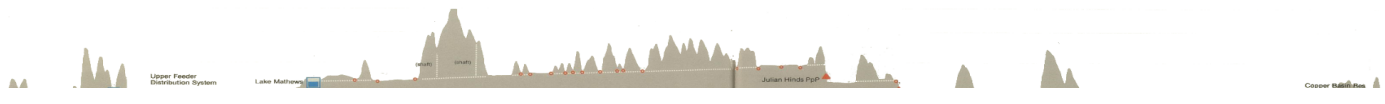
The SCUBA data platform developed for CaDC water managers automates the integration of water usage and contextual data from local water utilities collectively serving twenty three million Californians.[2] SCUBA utilizes industry standard Amazon Web Services for digital infrastructure and Apache Airflow, the open source automation package, to coordinate the routine extraction, transformation, loading and deployment of analytics.[3]

The word "platform" is often used loosely in water and other public data discussions, so here we would like to offer a definition so we can be explicit by what we mean when we say SCUBA is a platform. A platform is a technology upon which other technologies are based. For example, roads are a platform technology for cars or bicycles. Similarly SCUBA is a platform technology for CaDC analytics deployed for local water utilities[4] and streamlining the secure sharing of data for research.

SCUBA integrates and CaDC local water utility managers regularly utilize the state's water data including evapotranspiration data from the California Irrigation Management Information System (CIMIS), land use information compiled by DWR, reservoir levels in the California Data Exchange Center (CDEC) and many, many other state water datasets in routine analyses and open source analytics.

---

[1] See here for metadata summarizing data integrated into SCUBA: bit.ly/scuba_metadata_download
[2] See here for the CaDC onboarding protocol: https://docs.google.com/document/d/1ppWv3T2Sjr0zNP40PB5C_4NAREC1n_JtOvudpvZBNmE/edit?usp=sharing
[3] Apache Airflow was developed by Maxime Beauchemin at Airbnb after he developed several similar automation tools at other leading Silicon Valley companies. See here for a podcast on the origins of Airflow with Maxime: https://www.stitcher.com/podcast/astronomer-2/the-airflow-podcast/e/53193033
[4] See here for a demo portal: demo.californiadatacollaborative.com/smc/



CALIFORNIA DATA COLLABORATIVE

Note there are no local water utility managers listed in the AB1755 technical working group, and we see an opportunity to improve local water utility engagement in this important initiative. We at the CaDC thus would like to be active and constructive participants in the implementation of AB1755. To that end, we have a couple initial high level questions and comments on the preliminary progress report.

In particular, it may be unclear to many readers what is actually meant by the water data platform referenced throughout the document as a definition is never provided. In developing the data platform called for by AB1755, the draft progress report emphasizes the concept of data federation, which the CaDC agrees will be critical for successful implementation of AB1755:

> *"As a starting point, DWR and its partners aim to federate and populate the California Government Operations Agency and California Natural Resources Agency open data portals to allow users access to available State water and ecological datasets. In time, additional portals will join the federation, bringing users greater access to available water and ecological data."* [5]

Understanding that further detail will come in the Implementation Plan, it would be helpful to understand how that federation is proposed to be accomplished at a high level in the April Progress Report. How would for example a local water manager be able to provide input to quality control or answer user questions about their Urban Water Management Plan (UWMP) data that is federated across multiple portals?

More broadly, it would be helpful for this document to discuss how the proposed protocols and implementation strategy would work in terms of specific datasets like the local utility UWMPs. CaDC interns have conducted a preliminary inventory of datasets in the data sources listed in the AB1755 legislative text that may be useful to that end.[6] That detail would be helpful in enabling the virtuous iterative cycled described in the water data for decision-making summary blog post:

> *"Engagement between data system developers and end users is, ideally, an ongoing and iterative process."* [7]

In that spirit, we would like to discuss specific technical areas of overlap between ongoing CaDC water data projects and AB1755 implementation in the future and again hope DWR can present at our April 29th workshop. These opportunities include: 1) collaboratively developing a visualization of California's water supply, 2) sharing data protocols developed by the CaDC including the open water rate specification, and 3) discussing how to leverage open source technologies underlying SCUBA platform for AB1755 implementation.

Thank you for the opportunity to comment and we look forward to collaborating.

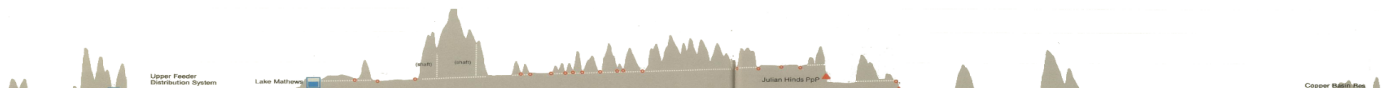
Patrick Atwater
On behalf of the California Data Collaborative

---

[5] See "AB 1755 Draft Progress Report": https://www.water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/All-Programs/AB-1755/AB1755ProgressReportinitialdraft13018v42.pdf?la=en&hash=0B9CB4C2BE515DDEBFCCFF1D9D436FD4707ED438

[6] See: https://docs.google.com/spreadsheets/d/13WHmztwBOjyCEuSNQd09nl2wIPXrAUbtqidpiNKR13k/edit#gid=1567667 11

[7] See "Water Data for Decision Making": https://www.law.berkeley.edu/research/clee/research/wheeler/data/



CALIFORNIA
DATA
COLLABORATIVE

In addition to core CaDC staff and partner data action team members, this work has been enabled by an inspiring team from across the globe. Special thanks to Varun Adibhatla (SCUBA platform adviser), Tony Castelletto (system architecture research fellow), Vyki Englert (civic technology best practices), Mike Amodeo (reservoir explorer graduate intern developer), Neil Verosh (Gittes AB 1755 data inventory), Vipassana Vijayarangan (Gittes AB 1755 data inventory) and the many CaDC interns who went above and beyond -- often volunteering their nights and weekends -- to provide research and prototyping services to help realize the dream of California's first statewide water data platform.